

SuperSegger: robust image segmentation, analysis and lineage tracking of bacterial cells

Stella Stylianidou,¹ Connor Brennan,¹
Silas B. Nissen,² Nathan J. Kuwada³ and
Paul A. Wiggins^{1,4,5*}

¹Departments of Physics, University of Washington,
Seattle, WA 98195, USA.

²StemPhys, Niels Bohr Institute, University of
Copenhagen, Copenhagen 2100, Denmark.

³Department of Physics, Central Washington
University, Ellensburg, WA 98926, USA.

⁴Department of Bioengineering, University of
Washington, Seattle, WA 98195, USA.

⁵Department of Microbiology, University of
Washington, Seattle, WA 98195, USA.

Summary

Many quantitative cell biology questions require fast yet reliable automated image segmentation to identify and link cells from frame-to-frame, and characterize the cell morphology and fluorescence. We present *SuperSegger*, an automated MATLAB-based image processing package well-suited to quantitative analysis of high-throughput live-cell fluorescence microscopy of bacterial cells. *SuperSegger* incorporates machine-learning algorithms to optimize cellular boundaries and automated error resolution to reliably link cells from frame-to-frame. Unlike existing packages, it can **reliably segment microcolonies with many cells, facilitating the analysis of cell-cycle dynamics in bacteria as well as cell-contact mediated phenomena.** This package has a range of built-in capabilities for characterizing bacterial cells, including the identification of cell division events, mother, daughter and neighbouring cells, and computing statistics on cellular fluorescence, the location and intensity of fluorescent foci. *SuperSegger* provides a variety of postprocessing data visualization tools for single cell and population level analysis, such as histograms, kymographs, frame mosaics, movies and consensus

images. Finally, we demonstrate the power of the package by analyzing lag phase growth with single cell resolution.

Introduction

Quantitative fluorescence imaging has become an important tool for the study of the bacterial cell biology (Teeffelen *et al.*, 2012). Most cellular processes from chromosome segregation to replication are stochastic in nature and exhibit significant cell-to-cell variation (Kuwada *et al.*, 2013; Kuwada *et al.*, 2015b). This variation makes the quantitative statistical analysis of a significant number of cells essential to understanding many biological processes at the single cell level (Kuwada *et al.*, 2015b). Although it is straightforward to image fields of view containing thousands of bacterial cells, the development of fast, reliable and automated methods for the quantitative analysis of this data, including segmentation (the identification of regions in each image corresponding to cells) and linking the cell regions from frame-to-frame, remains a significant computational challenge.

In this article, we present *SuperSegger*, an automated MATLAB-based image processing and analysis package. To date, we have successfully applied our package to a number of problems in the cell biology of bacterial cells, including a genome-scale characterization of protein localization dynamics in *Escherichia coli* throughout the cell cycle (Kuwada *et al.*, 2015a,b), detailed analysis of the function of the Type VI Secretion System in *Pseudomonas aeruginosa* (LeRoux *et al.*, 2012; Russell *et al.*, 2013; LeRoux *et al.*, 2015a), chromosome segregation in *E. coli* (Kuwada *et al.*, 2013; Cass *et al.*, 2016) as well as a number of other applications (O'Connor *et al.*, 2012; Stylianidou *et al.*, 2015). *SuperSegger* is particularly well suited for high-throughput analysis of cell-cycle dynamics of proteins and complexes by time-lapse fluorescence microscopy in single bacterial cells. Although the software is optimized for rod-like bacterial cells, it incorporates machine-learning algorithms, to optimize cell boundaries for other cell shapes. *SuperSegger* can follow a

Accepted 19 August, 2016. *For correspondence. E-mail pwiggins@uw.edu; Tel. 206 616 2618; Fax 206 685 0635.

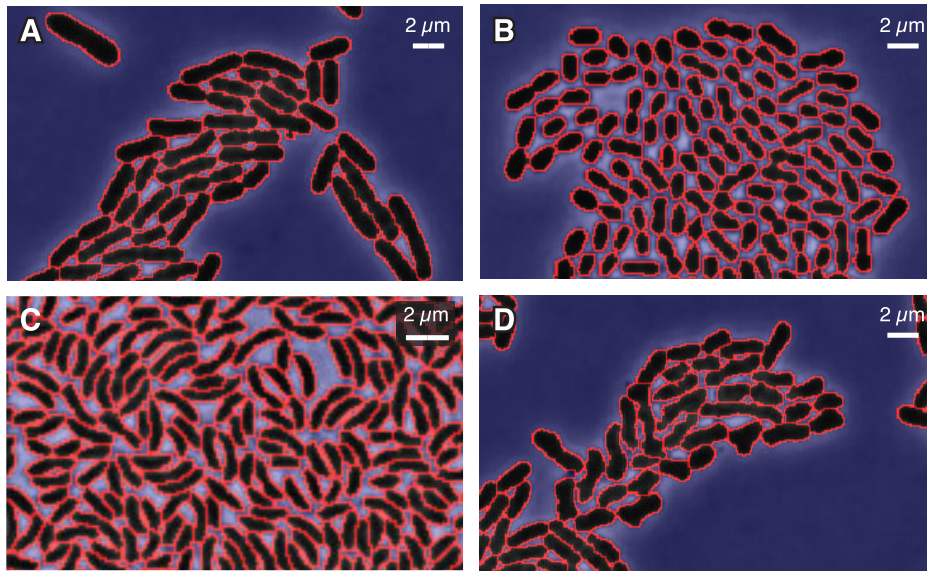


Fig. 1. Segmented brightfield images of bacteria. *SuperSegger* can successfully segment images of differently shaped cells and cells in close contact such as **Panel A:** *E. coli*, **Panel B:** *A. baylyi*, **Panel C:** *C. crescentus* and **Panel D:** irregularly shaped *E. coli* cells. Red are the computer generated boundaries.

cell lineage for many generations, identifying full cell cycles by linking cells from frame-to-frame in time-lapse imaging. The software performs a detailed characterization of each cell facilitating a wide range of analyses. Postprocessing tools are included for analysis at both the single cell and population level, such as fluorescence kymographs, frame mosaics, consensus images and plotting tools for a variety of cellular characteristics. Powerful tools for subpopulation analysis have also been developed that allow the user to gate (generate cell subpopulations) on a wide range of cellular characteristics. The software can be downloaded from the Wiggins Lab website where the user manual and documentation on all the methods can be found. It can then either be run at the command-line or through a graphical user interface (Figs. S1 and S2).

Although a number of segmentation software packages are already available (Guberman *et al.*, 2008; Christen *et al.*, 2010; Wang *et al.*, 2010; Young *et al.*, 2011; Paintdakhi *et al.*, 2016), our software was developed because we found the existing packages were not well suited to our own experiments and analysis. In our experience, these packages cannot track cells over multiple generations without significant error correction by hand. This approach is not practical when analyzing a large number of cells. In contrast we demonstrate that *SuperSegger* can segment multiple generations of cell division without error. Therefore, *SuperSegger* offers strengths complementary to these other packages, especially in experimental contexts dependent on time-lapse imaging and cultures where cells have many neighbours.

Results

High-throughput analysis

Our software is capable of segmenting high-throughput datasets with multiple time frames and xy positions. Furthermore, it takes advantage of parallel computation for each separate xy frame and time frame, decreasing the processing time. Datasets taken at high frame rates can expedite the segmentation process by taking advantage of the option to segment images every *k* frames. In this case, the skipped frames receive the boundaries of the cells from the last frame segmented in the sequence.

Reliable and flexible segmentation

The segmentation parameters can be optimized on different cell shapes making *SuperSegger* highly flexible. We have trained parameters on *E. coli*, *P. aeruginosa*, *Caulobacter crescentus* and *Acinetobacter baylyi*. In Fig. 1, we show examples of segmentation of *E. coli* (Panel A), *A. baylyi* (Panel B), and *C. crescentus* (Panel C) and *E. coli* with abnormal cell morphology (Panel D). In our experience, these segmentation parameters generalize well between cell types and experimental conditions. For instance, we use parameters trained on *E. coli* to segment *Bacillus subtilis*. Nevertheless, for different species, abnormal morphologies or different pixel sizes than the supplied (60 nm and 100 nm), the user can interactively produce a training set of identified true and false boundaries and generate new segmentation parameters. For more information on training a dataset see the Supporting Information. Also, *SuperSegger*

includes tools for interactively correcting boundaries after segmentation if needed.

Postprocessing tools

SuperSegger includes a postprocessing GUI to visualize the segmentation, fluorescence localization, focus position and tools to interactively modify incorrect cell boundaries (Fig. S2). Further, it provides a variety of analysis tools at the single cell level as well as at the population level.

At the single cell level, the user can produce through the postprocessing GUI a rotated and masked fluorescence profile for a cell for each frame during its lifetime, which we refer to as the cell tower (Fig. 6, Panel B). The fluorescence profile along the long axis of the cell can be displayed in a kymograph (Fig. 6 Panel C) which is the summed fluorescence along the short axis of the cell with respect to time. A movie for each cell can also be made (see Movie S1).

At the population level, our postprocessing GUI produces histograms for any of the features found in the Clist table (Fig. 6, Panel F), dot-plots of the Clist descriptors against each other (Fig. 6, Panel G) and plots for the descriptors in the Clist-in-time for all cells versus the time frame (Fig. 6, Panel H). In addition, by averaging over multiple single cell files, a quantity of interest for the population can be visualized, such as the mean cell-cycle-dependent fluorescence localization pattern, which we refer to as the consensus image (Fig. 6, Panel A). The consensus images are produced by computationally interpolating each single cell onto a reference cell cycle with regular shape, growth rate and lifetime followed by averaging the fluorescence profile (Kuwada *et al.*, 2015b).

Snapshot benchmark

A standard metric for the performance of a segmentation algorithm is looking at the segmentation error rate. Our software performs exceptionally well in this metric. In a snapshot image of 824 *E. coli* cells in microcolonies of about four cells, 99.4% of the 1198 boundaries were correctly classified resulting to 99.3% of the cells to be correctly segmented (Fig. S4). We believe that the measurement of error rate in a snapshot image is a misleading metric. The error rate in snapshot analysis depends sensitively on cell density since snapshot images of isolated bacterial cells can be segmented without error.

Time-lapse benchmark

What makes the bacterial cell segmentation problem challenging are the cell contacts and microcolonies. In

microcolonies, cells at the boundary with excellent contrast and cells in the interior of the microcolony with poorer contrast due to the shade-off artifact must be segmented simultaneously. Therefore, a natural test of the algorithm is the analysis of cell proliferation on an agarose pad starting with a single isolated cell. Although the culture starts with isolated cells, a microcolony forms and grows for roughly seven rounds of division before the culture begins to become multilayered.

To test the performance of *SuperSegger* in segmenting microcolonies, we analyzed 80 initially isolated cells grown on M9 agarose pads for 420 min. (See Supporting Information.) In each dataset the software marked the first segmentation error, defined as a frame-to-frame substantial area change or loss of a cell lineage. The results are shown in Fig. 2.

In Panel A, we show an example of a segmented dataset with 363 cells in the final frame and 727 cells segmented cumulatively. (Cells refer to the number of actual cells observed during the experiment, not regions segmented in a single frame.) In this dataset, 543 cells were segmented before the first error occurred and, in the final frame, only one cell in the middle of the colony was segmented with an error. (See all frames in Movie S2.) To understand why the segmentation problem is challenging, consider that in order to correctly segment this microcolony, 30,179 boundaries needed to be classified during the time course.

In these 80 datasets, the mean number of cells segmented before an error was 174 cells, and 30% of the datasets were segmented without error (Fig. 2, Panel B). One of the main mechanisms resulting in segmentation errors is the overlap of cells that occurs when the colony begins to become multilayered. In this experiment, cells began to overlap at about 400 min and a corresponding rise in the error rate is clearly seen in Panel C. Even with this increase, the error rate remains low: At the end of the experiment only 2.5% of the cells have segmentation errors.

Single-cell analysis of lag phase

To demonstrate the power of *SuperSegger* for quantitative analysis, we analyzed cell proliferation during lag phase in single bacterial cells. Lag phase refers to the period during which bacteria adapt to changes in growth conditions. It usually refers to the transition that occurs when a stationary-phase culture is reinoculated into fresh media, resulting in log phase growth after a brief period of slow growth (Madigan *et al.*, 2014). Although lag phase growth has been characterized and analyzed at a population level (Rolfe *et al.*, 2011), little is known about the nature of lag phase on a single cell level

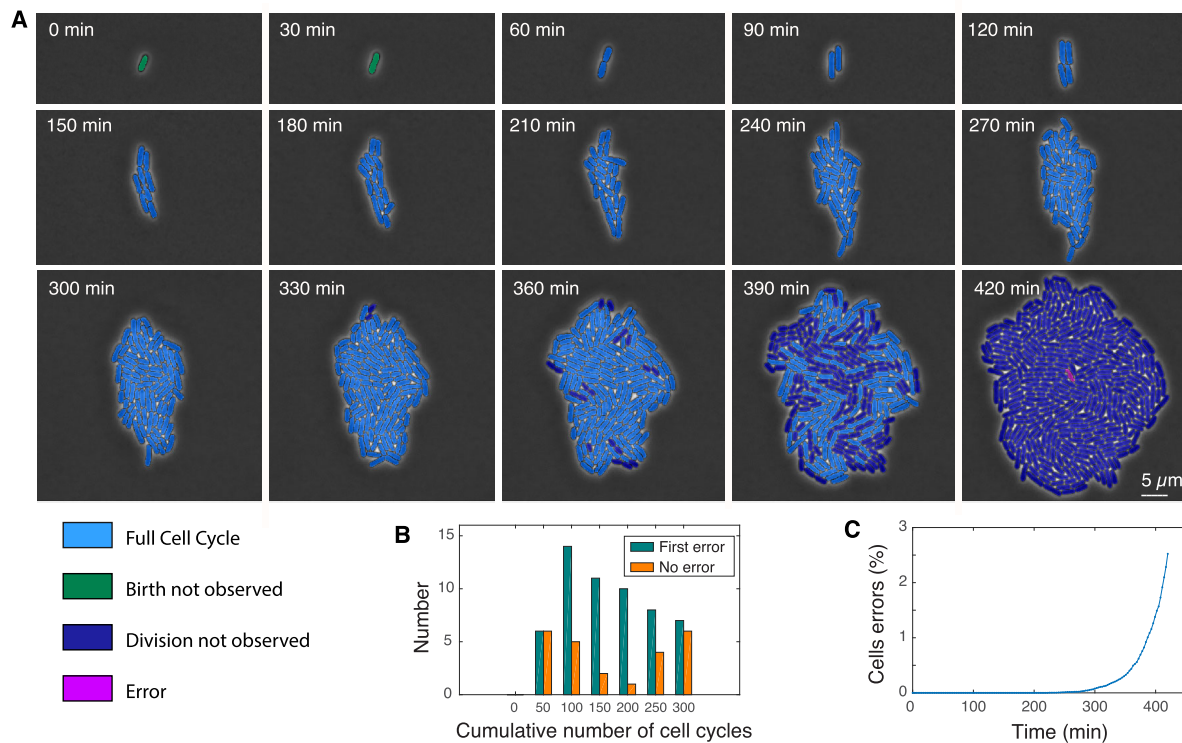


Fig. 2. Segmentation of proliferation from single *E. coli* cells.

Panel A: Frame mosaic of segmented images of a microcolony. A cell is marked with an error at the end of the time-lapse. (Only 15 of 141 frames are shown. See Movie S2 to view all frames.)

Panel B: A histogram of the cumulative number of cell cycles observed before the first error or no error occurred at each dataset ($N = 80$).

Panel C: The percentage of segmented cells with errors with time for all datasets ($N = 80$).

(Levin-Reisman *et al.*, 2010; Madar *et al.*, 2013; Gefen *et al.*, 2014; Jöers and Tenson, 2016). In particular, it is unclear whether the observed delay associated with lag phase is the result of slow growth of all cells in the population or a graded response with range of growth rates at the cellular level. The observation of significant cell-to-cell variation could be evidence of a *bet-hedging* strategy since the aberrant transition of cells from a dormant to a growing state would reduce cell fitness in a fluctuating environment (Balaban *et al.*, 2004; Veening *et al.*, 2008; Jöers and Tenson, 2016).

To investigate the nature of lag phase with single cell resolution, we analyzed cell proliferation upon re-inoculation on an agarose pad. We characterized progenitor cells in both log and late-stationary/death phase growth. (See Supporting Information.) The analysis is shown in Fig. 3. Panels A–C show the proliferation of a single cell over 5 hours to a final population of 115 cells. The microcolony was segmented without error. Panel A shows the lineage tree for the microcolony and Panel B shows a frame mosaic of the colony. Panel C shows the growth curve of the microcolony. Divisions early in growth are closely synchronized, resulting in a

characteristic step-like increase in the number of cells. As subsequent divisions lead to a de-synchronization, steps later in growth are smoothed. In the limit of large time, the growth curve approaches a straight line (on a semi-log plot), consistent with the exponential growth of an unsynchronized population.

The microcolony growth curves for log ($N = 82$) and stationary phase ($N = 119$) progenitor cells are shown in Panel D. Over 70,000 microcolony images were segmented in the analysis. The log (blue) and stationary phase (yellow) cells both proliferate with nearly identical growth rates, as is clear from the slopes of the growth curves. The principle difference between the curves is a delay in lag phase for stationary phase cells, as is expected from population-level analysis of lag phase after stationary phase (Madigan *et al.*, 2014). For a number of stationary phase cells the first division event (1→2) is observed more than 120 min after the cells were inoculated.

To quantify this delay and analyze the persistence of slow growth after the first division, we plot histograms for the duration of the cell cycle for the first five generations of growth (Fig. 3, Panel E). The 0* generation

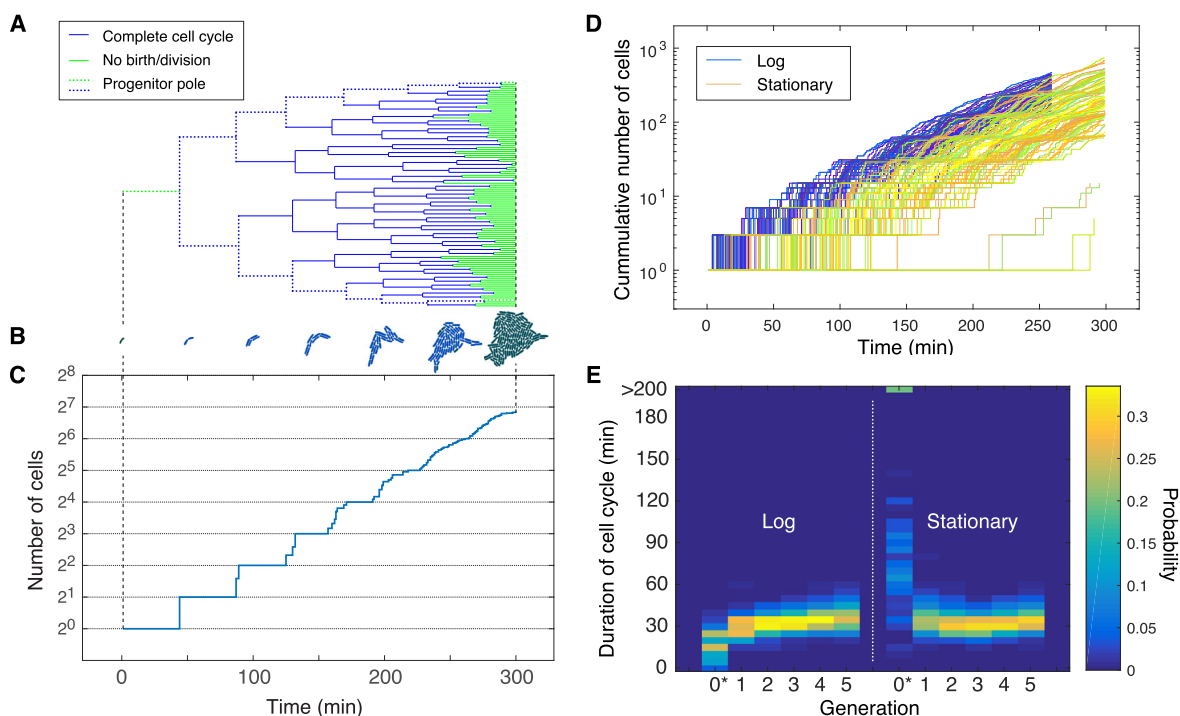


Fig. 3. Panel A: Cell lineage tree for a progenitor single cell.

Panel B: Frame mosaic of microcolony. Blue cells are observed for a complete cell cycle.

Panel C: Growth curve for a single microcolony.

Panel D: Growth curves for progenitor cells in either log (blue, $N = 82$) or stationary phase (yellow, $N = 119$).

Panel E: Duration of cell cycle by cell generation of growth on the agarose pad. The cell generation is shown on the x-axis for the two populations (log and stationary phase). The 0* generation is defined as the incomplete cell cycle of the initial progenitor cell.

represents the initial incomplete cell cycle of the progenitor cell. Since the cells do not begin the experiment synchronized, the 0* duration is not directly comparable to the duration of subsequent complete cell cycles. As expected, in generation 0* the log-phase progenitor cells execute an incomplete cell cycle in a period shorter than later complete cell cycles (30 min). In contrast, the stationary-phase progenitor cells have a much longer duration 0* generation time. About 80% of the cells divide in the first 200 min and show a wide distribution of cell cycle durations, centered around an hour. Nearly 20% of cells do not divide in the first 200 min. It is important to note that not all these cells are dead: 13% of these cells divide between 200 min and 300 min (when the time course ends).

After the 0* generation, we do not see a strong memory effect of the duration of generation 0 on the duration of the next generation, generation 1 ($p = 0.07$) (Fig. S5). The first full cell cycle observed, generation 1, shows a slight shift relative to subsequent cell cycles. (Log phase cells also show a small lag phase due to changes in growth conditions on the agarose pad.) But our experiments demonstrate that this effect is small and slow growth does not appear to persist.

Discussion

Comparison with existing software

We compared the performance of *SuperSegger* with existing software packages. All packages we tested could successfully segment cells that were well separated or formed small colonies of a couple cells, but since many of our applications demand full cell cycle imaging over multiple generations, it is necessary to reliably segment microcolonies containing many cells. Segmenting cells in a microcolony is particularly difficult because contrast is lost towards the center of the microcolony. In our tests of performance, *SuperSegger* outperformed the other software packages we tried (Young *et al.*, 2011; Paintdakhi *et al.*, 2016) in segmenting these microcolonies. Other packages either incorrectly merged or split multiple cells in the interior of the microcolony, even in a single frame (Fig. S6). Lineage tracking requires the linking of hundreds of error-free segmented frames.

Generalization

We note that one of the strengths of *SuperSegger* is that it does not require users to fine-tune the segmentation

parameters from application to application and does not require retraining. Our reported benchmarks are measurements of generalization error, not training error. The segmentation parameters were trained against a different strain of *E. coli*, grown under different conditions. No dataset specific optimization of the parameters was required to perform the segmentation. Other packages often require either setting specific parameters for each dataset or the user to correct the boundaries by hand after segmentation. Although as we mentioned before *SuperSegger* includes an interactive tool for boundary correction, we have not needed this feature since errors are rare enough that cells with segmentation errors can be discarded.

A model for lag phase growth

The order of magnitude increase in the accuracy of segmentation using *SuperSegger* facilitates the analysis of microcolonies of stationary and log phase progenitors with single cell resolution and reveals new subtleties in lag phase growth phenomenology. Our experiments show a wide distribution in the delay before the first division with 20% of cells exhibiting an extremely slow transition to the growing state ($\tau > 200$ min). As noted above, not all these cells are dead and some transition after 200 min, but before the end of the experiment at 300 min. These observations suggest a qualitative model for stationary and death phase with three cell states: (i) stationary, (ii) stationary-dormant and (iii) dead cells. The stationary-dormant cells are those that we have identified which have a significant delay before the first cell division. It is possible that this state is an intermediate state between stationary phase and cell death. Alternatively, this state may be the result of a bet-hedging strategy where a small subset of the population of cells does not transition into the growing state. This may insure against short-lived increases in nutrients which might lead cells to transition into the growing state when the optimal behaviour is to remain dormant (Veening *et al.*, 2008).

Conclusion

SuperSegger is a powerful automated image processing and analysis package, well suited for high-throughput time-lapse fluorescence microscopy of cells (LeRoux *et al.*, 2012; O'Connor *et al.*, 2012; Kuwada *et al.*, 2013; Russell *et al.*, 2013; Garmendia-Torres *et al.*, 2014; Javer *et al.*, 2014; Kuwada *et al.*, 2015a,b; Lampo *et al.*, 2015; LeRoux *et al.*, 2015a; Stylianidou *et al.*, 2015; Cass *et al.*, 2016). It provides reliable and flexible segmentation which can be trained to optimize performance in a wide variety of experimental contexts. The multiple data structures output by *SuperSegger* facilitate its use

in many experimental imaging contexts. The faithful segmentation and tracking of bacterial cells with neighbours facilitates its use of quantitative imaging-based analysis in new problems, like bacterial cell-contact mediated phenomena (LeRoux *et al.*, 2012; Russell *et al.*, 2013; LeRoux *et al.*, 2015a,b), where cell contacts are essential and applications where investigators want to study cell proliferation over multiple generations.

Experimental procedures

Overview

SuperSegger automates segmentation, linking and cytometry analysis and requires no user input (Fig. 4). In short: The user specifies the data directory and selects the appropriate segmentation parameters (typically using a set supplied with the package). *SuperSegger* loads images for each time point, aligns the images, identifies the cell regions from the background, tracks each cell region from frame-to-frame by linking corresponding regions and calculates fluorescence and structural characteristics for each cell in each frame. *SuperSegger* generates three different types of outputs: Frame files, Clist matrices and Cell files, which are further described below.

Image alignment

During long time-lapse imaging, stage drift often results in the accumulation of offset of the field of view between the initial and the final frames. In our own experience, this problem is often aggravated by imaging multiple xy positions during a time course. To implement frame alignment, we use a fast and memory-efficient algorithm by (Guizar-Sicairos *et al.*, 2008) which determines the frame alignment to subpixel resolution using crosscorrelation.

SuperSegger includes a number of important alignment features: (i) Alignment can be performed in any master channel (phase contrast, fluorescence, etc). (ii) A constant relative offset can be applied to each channel to compensate for known offsets between channels. (iii) Alignment can be performed against the previous frame, suitable for long time-lapse experiments, or against the first frame, suitable for short experiments where no significant cell growth occurs. (iv) Frames with high alignment error or out of focus are skipped automatically. (v) The image is preserved without cropping, facilitating the analysis of microcolonies that drift in (or out) of the field of view during long time courses.

Cell segmentation

The process of identifying the regions of the image which correspond to cells is called image segmentation. We developed the segmentation algorithm for use on phase contrast images (Fig. 5, Panel A), although the algorithm also performs well on inverted fluorescence images of cells with cytoplasmic fluorescence (J. Russell, pers. comm.).

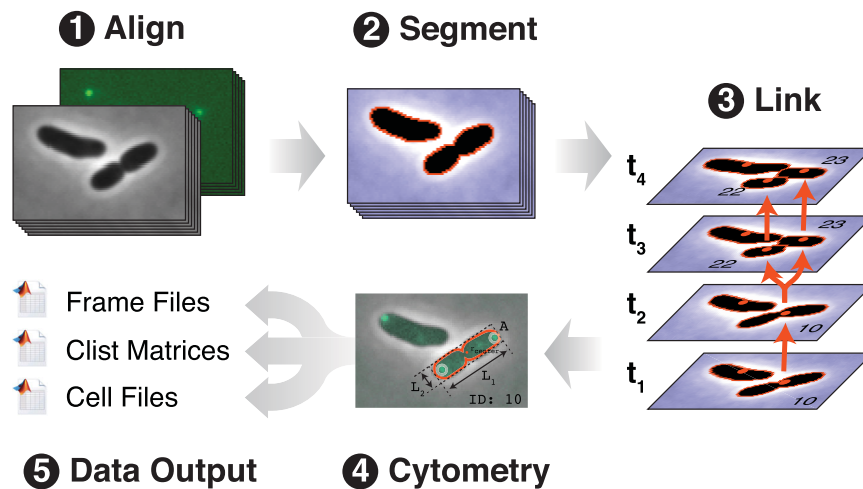


Fig. 4. General work flow in *SuperSegger*. The fluorescence and phase images are processed and aligned.

During segmentation the cell regions are identified from the background. Then each cell region is linked to a corresponding cell region in the next frame and the cells receive unique ID numbers. Next, the properties and fluorescence characteristics of each cell are calculated. Finally, the program outputs three different types of outputs: Frame files, Clist matrices and Cell files.

The first step of segmenting cells is differentiating cell microcolonies (or clusters) from empty regions of the frame. This initial microcolony mask is created by a thresholding operation on the phase contrast image which identifies clusters by their low relative intensity (Fig. 5, Panel B). The threshold value is set as a constant in the segmentation

parameters. Since the threshold is applied after the image intensity is normalized, it does not need to be adjusted from dataset to dataset. The phase contrast shade-off (the loss of phase-based contrast in the middle of microcolonies) and halo artifacts complicate the interpretation of the phase image and the image is filtered to compensate for

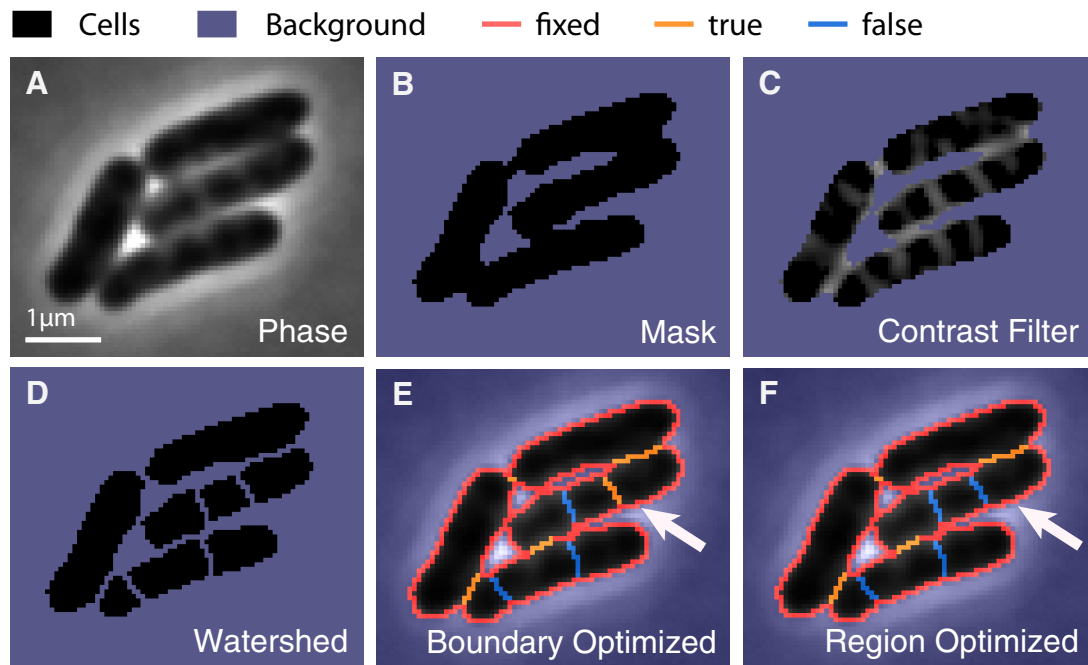


Fig. 5. Image segmentation procedure in *SuperSegger*. *E. coli* cells are shown as a representative example.

Panel A: Original phase image.

Panel B: Mask of cells using intensity thresholding.

Panel C: Phase image after the contrast (maximum principal curvature) filter and the mask are applied.

Panel D: Boundaries found using the watershed function.

Panel E: Boundaries after the boundary optimization and

Panel F: Boundaries after region optimization. Boundaries are divided into fixed boundaries (red), boundaries that the software classified as true (orange) and false (blue). In Panel E on the right side, there is a boundary incorrectly set as true after boundary optimization which is set to false after region optimization in Panel F.

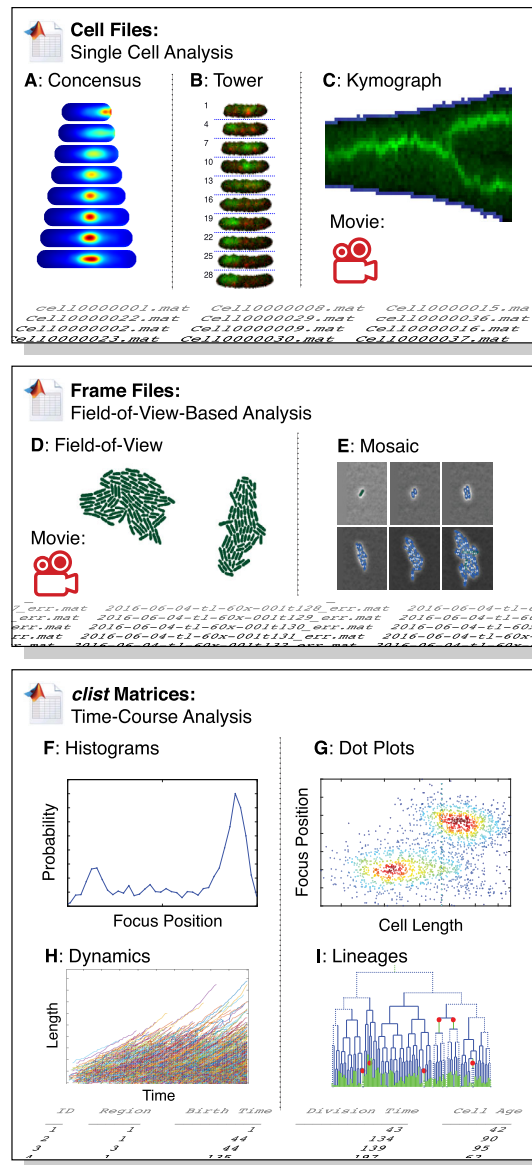


Fig. 6. SuperSegger output and analysis.

Cell files contain all time points for a single cell. Cell data can be visualized as: **Panel A:** Consensus images, the mean cell-cycle-dependent fluorescence localization pattern [example of Mall, a transcriptional repressor (Kuwada *et al.*, 2015a)]. **Panel B:** Cell towers (example of mini-F plasmids labeled with mCherry-TetR and ParA-GFP (ATPase) [Jetswaart *et al.*, 2014]). **Panel C:** Kymographs (example shows of *parS* cassette inserted at *oriC* and labeled with GFP-ParB (Cass *et al.*, 2016) and cell movies (Movie S1). **Frame files** contain the data from a single field of view at a time point. Frame data can be visualized as: **Panel D:** Field-of-view images, **Panel E:** Frame mosaics, and movies (Movie S2). **The clist matrices** contain a summary data for all cells at all time points. Clist data can be visualized as: **Panel F:** Histograms (example of probability density of position of mRNA-MS2 foci [Stylianidou *et al.*, 2015]), **Panel G:** Dot Plots (example of focus separation of replication forks versus cell length (S. Mangiameli, unpublished), **Panel H:** Cell Dynamics plots [example of cell length with time (Stylianidou *et al.*, 2015)] and **Panel I:** Cell Lineages.

these artifacts (Inoué and Spring, 1997) (Fig. 5, Panel C). The outline of the mask of each microcolony is set as a fixed cell boundary during the rest of the segmentation procedure.

To identify cell boundaries within a microcolony, we use the watershed operation. The watershed operation identifies the catchment basins of local intensity minima, which are the dark regions of the image corresponding to cells. This

results to over-segmentation: the typical cell is subdivided into multiple watershed regions (Fig. 5, Panel D). Watershed regions too large to represent cells are further subdivided to ensure that each cell is represented by at least one watershed region.

The cell regions are formed by merging watershed regions. This is done by classifying the state of boundaries between them (Fig. 5, Panel E). The boundaries are

classified as true or false using a neural network trained on the boundaries of the specific cell type, growth conditions, pixel size, and so forth. The input variables for each boundary include: mean and minimum intensity of the boundary, the second derivative of the intensity of the boundary, the length of the boundary, the area of neighbouring regions, axes of neighbouring regions, and so forth.

The state of boundaries neighbouring low scored regions are further informed by globally optimizing a score function which includes the scores of both regions and boundaries in question (Fig. 5, Panel F). To score a cell region, we use a second neural network trained to classify true or false regions. The input variables for each cell region include: the length of the long and short axis of the region, the area of the region, the neck of the region, and so forth. Depending on the number of marginal boundary state classifications, either an exhaustive or stochastic (simulated anneal) search algorithm is used to identify the optimal segmentation.

Region linking

In snapshot analysis, the cell segmentation process is now complete, but for time-lapse analysis, the cell regions in successive frames must be linked to generate a cell cycle. The linking algorithm links each region to a single region (or a pair of regions) in the successive frame. A linking cost function is computed from (i) the spatial overlap between regions, (ii) the distance between the region centroids of the two regions and (iii) the change in the areas of the regions. The linking corresponding to the minimum cost is initially assigned, and the assigned regions are removed from the possible assignments until no possible assignment remains. The linking information from frame-to-frame is used to identify boundaries that may have been incorrectly classified as true or false. In the event that the area change is above a defined threshold the region is marked with an error flag. This makes it trivial to exclude incorrectly segmented cells at the end analysis.

Cell cytometry

After the segmentation process is complete, a detailed and time-dependent cell cytometry analysis is performed. The cell descriptors are stored in a cell data structure. Each distinct cell is given a unique ID. The cell descriptors include cropped phase and fluorescence images of the cell, the length and width of the cell, the major and minor principal axis, the area, the location of fluorescent foci in global and local coordinates, the IDs of daughters, sister, mother and neighbour cells, the frame of death and birth of the cell, and so forth (Fig. S3).

Analysis of fluorescence channels

SuperSegger includes a number of generic tools for the quantitation of cellular fluorescence, including the computation of statistics on fluorescence intensity and the identification of punctate foci. The mean background fluorescence

intensity for each frame is subtracted from the fluorescence images. To identify and score foci, *SuperSegger* uses an image-curvature method which is specifically engineered to avoid the identification of false positive foci due to background intensity from cytoplasmic fluorescence (C. Brennan, unpubl. data). Foci detection is performed in the union of all cell regions and foci are then allocated to cells to circumvent the double counting of foci close to the membrane at the interface between cells (LeRoux *et al.*, 2012; Russell *et al.*, 2013; LeRoux *et al.*, 2015a, b).

Complete cell cycle imaging

In order to study the dependence of fluorescence localization with the cell cycle phase or the new or old pole of the cell, it is important to identify complete cell cycles. *SuperSegger* identifies division and birth events. After a division is observed the program identifies the new poles, created at the last cell division, or old, created by an earlier division. During the identification of division events, the mother and daughter cells are also tracked. Starting at any frame, it is possible to track the lineage of a single cell both backwards and forwards in time and, for example, to investigate fluorescence localization through the generations.

Data output

Due to the variety of experimental demands and the large size of the typical dataset, *SuperSegger* slices the data into three output formats: The frame files organize the data by frame and contain all the information about all cells in a single time point. The cell files organize the data by cell and contains all data for a single cell over all time points (Table S2–S7, Fig. S3). Finally, the Clist file contains (i) a matrix of over 70 cellular descriptors (including cell age at division, cell length at birth and division, average fluorescence intensity, and so forth [Table S1]), organized by cell ID, and (ii) Clist-in-time structure with several cellular descriptors (cell length, average fluorescence intensity) for each time frame organized by cell ID (Fig. 6).

Gating functionality

SuperSegger implements a powerful gating functionality, inspired by flow cytometry, which facilitates the selection of subpopulations of cells that match specific criteria described using the Clist cellular descriptors. Gating significantly streamlines many common analyses. For instance, it is straightforward to limit analysis to cells which are (i) segmented without error, (ii) observed through a complete cell cycle and (iii) have an average fluorescence intensity above a specified cutoff. We describe this powerful approach to quantitative cell biology in more detail elsewhere (J.A. Cass, unpubl. data).

Acknowledgements

We would like to thank the members of the Wiggins Lab for testing *SuperSegger* during the development process, K.

Cheveralls for helping with initial development of the software, P.B. Benjamin and Z. Gitai for the images of *C. crescentus* cells, J.A. Cass for the *oriC* Kymograph image and the *A. baylyi* images, and finally S. Mangiameli for the dot plot image. This work was supported by University of Washington Royalty Research Fund, Sloan-BR2011-110, NSF-PHY-084845, NSF-MCB-1151043-CAREER and the Danish National Research Foundation.

Author Contributions

SS: software; validation; visualization; acquisition of data; methodology; writing; CB: software; SB: software; NJK: software; PAW: software; validation; visualization; acquisition of data; methodology; writing; supervision.

References

- Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004) Bacterial persistence as a phenotypic switch. *Science* **305**: 1622–1625.
- Cass, J.A., Kuwada, N.J., Traxler, B., and Wiggins, P.A. (2016) *Escherichia coli* chromosomal loci segregate from midcell with universal dynamics. *Biophys J* **110**: 2597–2609.
- Christen, B., Fero, M.J., Hillson, N.J., Bowman, G., Hong, S.-H., Shapiro, L., and McAdams, H.H. (2010) High-throughput identification of protein localization dependency networks. *Proc Natl Acad Sci USA* **107**: 4681–4686.
- Garmendia-Torres, C., Skupin, A., Michael, S.A., Ruusuvoori, P., Kuwada, N.J., Falconnet, D., *et al.* (2014) Unidirectional P-body transport during the yeast cell cycle. *PLoS One* **9**: e99428.
- Gefen, O., Fridman, O., Ronin, I., and Balaban, N.Q. (2014) Direct observation of single stationary-phase bacteria reveals a surprisingly long period of constant protein production activity. *Proc Natl Acad Sci USA* **111**: 556–561.
- Guberman, J.M., Fay, A., Dworkin, J., Wingreen, N.S., and Gitai, Z. (2008) PSICIC: noise and asymmetry in bacterial division revealed by computational image analysis at sub-pixel resolution. *PLoS Comput Biol* **4**: e1000233.
- Guizar-Sicairos, M., Thurman, S.T., and Fienup, J.R. (2008) Efficient subpixel image registration algorithms. *Opt Lett* **33**: 156.
- Ietswaart, R., Szardenings, F., Gerdes, K., and Howard, M. (2014) Competing ParA structures space bacterial plasmids equally over the nucleoid. *PLoS Comput Biol* **10**: e1004009.
- Inoué, S., and Spring, K.R. (1997) *Video Microscopy: The Fundamentals*. US: Springer, New York, USA, pp. 71–75.
- Javer, A., Kuwada, N.J., Zhicheng, L., Benza, V.G., Dorfman, K.D., Wiggins, P.A., *et al.* (2014) Persistent super-diffusive motion of *Escherichia coli* chromosomal loci. *Nat Commun* **5**: 3854.
- Jöers, A., and Tenson, T. (2016) Growth resumption from stationary phase reveals memory in *Escherichia coli* cultures. *Sci Rep* **6**: 24055.
- Kuwada, N.J., Cheveralls, K.C., Traxler, B., and Wiggins, P.A. (2013) Mapping the driving forces of chromosome structure and segregation in *Escherichia coli*. *Nucleic Acids Res* **41**: 7370–7377.
- Kuwada, N.J., Traxler, B., and Wiggins, P.A. (2015a) High-throughput cell-cycle imaging opens new doors for discovery. *Curr Genet* **61**: 513–516.
- Kuwada, N.J., Traxler, B., and Wiggins, P.A. (2015b) Genome-scale quantitative characterization of bacterial protein localization dynamics throughout the cell cycle. *Mol Microbiol* **95**: 64–79.
- Lampo, T.J., Kuwada, N.J., Wiggins, P.A., and Spakowitz, A.J. (2015) Physical modeling of chromosome segregation in *Escherichia coli* reveals impact of force and DNA relaxation. *Biophys J* **108**: 146–153.
- LeRoux, M., De Leon, J.A., Kuwada, N.J., Russell, A.B., Pinto-Santini, D., Hood, R.D., *et al.* (2012) Quantitative single-cell characterization of bacterial interactions reveals type VI secretion is a double-edged sword. *Proc Natl Acad Sci USA* **109**: 19804–19809.
- LeRoux, M., Kirkpatrick, R.L., Montauti, E.I., Tran, B.Q., Peterson, S.B., Harding, B.N., *et al.* (2015a) Kin cell lysis is a danger signal that activates antibacterial pathways of *Pseudomonas aeruginosa*. *Elife* **4**: e05701. <http://dx.doi.org/10.7554/elife.05701>.
- LeRoux, M., Peterson, S.B., and Mougous, J. (2015b) Bacterial danger sensing. *J Mol Biol* **427**: 3744–3753.
- Levin-Reisman, I., Gefen, O., Fridman, O., Ronin, I., Shwa, D., Sheftel, H., and Balaban, N.Q. (2010) Automated imaging with ScanLag reveals previously undetectable bacterial growth phenotypes. *Nat Methods* **7**: 737–739.
- Madar, D., Dekel, E., Bren, A., Zimmer, A., Porat, Z., and Alon, U. (2013) Promoter activity dynamics in the lag phase of *Escherichia coli*. *BMC Syst Biol* **7**: 136.
- Madigan, M.T., Bender, K.S., Buckley, D.H., Martinko, J.M., and Stahl, D.A. (2014) *Brock Biology of Microorganisms*. Pearson: San Francisco, USA, p. 149.
- O'Connor, J.R., Kuwada, N.J., Huangyutitham, V., Wiggins, P.A., and Harwood, C.S. (2012) Surface sensing and lateral subcellular localization of WspA, the receptor in a chemosensory-like system leading to c-di-GMP production. *Mol Microbiol* **86**: 720–729.
- Paintdakhi, A., Parry, B., Campos, M., Irnov, I., Elf, J., Surovtsev, I., and Jacobs-Wagner, C. (2016) Oufiti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. *Mol Microbiol* **99**: 767–777.
- Rolfe, M.D., Rice, C.J., Lucchini, S., Pin, C., Thompson, A., Cameron, A.D.S., *et al.* (2011) Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol* **194**: 686–701.
- Russell, A.B., LeRoux, M., Hathazi, K., Agnello, D.M., Ishikawa, T., Wiggins, P.A., *et al.* (2013) Diverse type VI secretion phospholipases are functionally plastic antibacterial effectors. *Nature* **496**: 508–512.
- Stylianidou, S., Kuwada, N.J., and Wiggins, P.A. (2015) The bacterial nucleoid drives cytoplasmic dynamics. *Biophys J* **108**: 541a.
- Teeffelen, S.V., Shaevitz, J.W., and Gitai, Z. (2012) Image analysis in fluorescence microscopy: bacterial dynamics as a case study. *Bioessays* **34**: 427–436.

- Veening, J.W., Smits, W.K., and Kuipers, O.P. (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **62**: 193–210.
- Wang, Q., Niemi, J., Tan, C.M., You, L., and West, M. (2010) Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. *Cytometry A* **77**: 101–110.
- Young, J.W., Locke, J.C.W., Altinok, A., Rosenfeld, N., Bacarian, T., Swain, P.S., *et al.* (2011) Measuring single-

cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat Protoc* **7**: 80–88.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.